1-1-2016

Selection Effect In Medigap Insurance Market With Multi-Dimensional Private Information

Yang Liu
Wayne State University,

SELECTION EFFECT IN MEDIGAP INSURANCE MARKET WITH MULTI-DIMENSIONAL PRIVATE INFORMATION

by

YANG LIU

DISSERTATION

Submitted To Graduate School

of Wayne State University

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2016

MAJOR: ECONOMICS

Approved By:

_____

Advisor Date

_____

_____

_____

DEDICATION

To my lovely mother

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my Ph.D. advisor Professor Allen Goodman for the tremendous support of my study and related research. Without his patience, motivation, and immense knowledge, I definitely couldn't finish this dissertation. His guidance helped me in all the time of research and writing of this dissertation. I could not have imagined having a better advisor and mentor for my Ph.D. study.

Besides my advisor, I would like to thank the rest of my dissertation committee, Professor Gail Summers, Professor Janet Hankin and Professor Stephen Spurr not only for their insightful comments, continuous encouragement and valuable suggestions, but also for the hard questions which incented me to widen my research from various perspectives.

Last but not the least, I use this opportunity to thank my family, my parents and my dear husband for supporting me spiritually throughout writing this dissertation and my life in general. Without their precious support it would not be possible to conduct my Ph.D. research and complete it finally.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Background

The total expenditure on healthcare and the cost as a percentage of GDP have increased steadily for years in most of countries, e.g., U.S. healthcare costs per capita increased from $4,878 in 2000 to $9,523 in 2014, and from 13.8% in 2000 to 17.5% in 2014 as the percentage of GDP and total amount is $3 trillion. National Health Expenditure-GDP ratio for developed countries like Germany and Switzerland is around 11%. Medicare spending is $618.7 billion in 2014, it accounts for 20 percent of total National Health Expenditure and grows by 5.5% compared to 2013 (NHE Fact Sheet, Center of Medicare and Medicaid Services).

Medicare

Medicare is a public health care program, which is supported by the Federal government for people aged 65 years or above and certain people aged under 65 years with disabilities. It plays a major role in providing health care service to the beneficiaries, and includes Medicare Part A, Part B, Part C and Part D.

Part A is the hospital insurance, which covers inpatient care in hospitals, including critical access hospitals and skilled nursing facilities. But long-term care is not included in Part A. Part B is the medical insurance, which covers outpatient care. Part C plans, which are known as Medicare Advantage plans, are offered through approved private insurance companies by Medicare. The newest addition to the Medicare family, Part D, is the prescription drug plan, which is optional and available to people enrolled in Parts A, B and most of Part C plans.

For an individual or his/her spouse who paid Medicare tax for at least 40 quarters (10 years) while working, there is no premium for Part A. Most of people need to pay premium if they are enrolled in Part B and Part D.

Part C (Medicare Advantage) plans are provided by private insurance companies. Medicare Advantage plans offer benefits of Part A and Part B and most of the Medicare Advantage plans also offer prescription drug coverage. Individuals can choose from HMO (Health Maintenance Organization), PPO (Preferred Provider Organization), Private FFS (Fee-for-Service), Special Needs Plan, and Medicare MSA (Medical Saving Account).

Medicare is not a comprehensive coverage and there is substantial amount of expenses, which are not compensated. Beneficiaries could choose supplemental insurance to cover the some of the "gaps" between total cost and reimbursed cost by Medicare.

Medigap

Medigap is also known as Medicare Supplemental Insurance, which is provided by private insurance companies used to cover some of the cost sharing required by original Medicare plan, for instance, deductibles, coinsurance, and copayments when using original Medicare.

Starting from 1992 the policies provided by insurance companies are standardized, and using single letters to represent various plans. Some plans only provide basic coverage, and some provide more advanced coverage. Insurance companies do not necessarily need to provide all possible plans. Medigap plan A is an necessity if insurance companies sell Medigap coverage and other advanced plans are optional.

Medigap market is highly regulated by the Federal government. Government restricts insurance companies to use pre-existing condition for risk classification during open enrollment period. Insurance companies have very limited information to use for the underwriting process. Therefore, strict regulation leads to severe information asymmetry and Medigap market would be an appropriate field to investigate economic problems caused by asymmetric information.

Overall, individuals can have two general options. One is to choose Part A plus Part B (optional) plus Part D (optional) plus Medigap coverage (optional). The other one is a Medicare Advantage plan. Any insurance company selling Medigap coverage to those who enrolled in Medicare Advantage plan is illegal.

1.2 Why Choosing Medigap Market to Examine Selection Effect

In a market with mandatory coverage, it is not allowed for people to step out of the market. Therefore there is no way to observe individuals' selection effect. Thus, a voluntarily chosen insurance is one basic requirement for investigating selection effect.

In voluntary insurance market with symmetric information, insurance companies would classify individuals and charge different premiums according to their classification outcome. Under this circumstance, there is no selection effect. In voluntary insurance market without government regulation and with asymmetric information, insurance companies take advantage of information, which is used to classify individuals. Meanwhile, individuals can use some private information that cannot be observed by insurance companies to make choice of insurance coverage. In this scenario, selection effect by

individuals and the screening by insurance companies are mingled. Thus, it is less likely to evaluate the selection effect.

However, when government intervenes insurance market and restricts the information that insurance companies can use, this problem is simplified since screening by the insurance companies are avoided and behaviors from demand side are isolated. Medigap insurance market falls into this category.

Firstly, Medigap is optional to those who are covered by Medicare Part A and Part B. Secondly, government intervention limits the information, which can be used by the insurance companies. In other words, information in Medigap insurance market is asymmetric.

1.3 Economic Consequences of Asymmetric Information in Medigap Insurance Market

Main consequence of asymmetric information is the market failure. Perfect symmetric information is the ideal case and unrealistic. Usually, it is hard for buyers and seller to know each other perfectly. When there are restrictions in insurance underwritings, it makes more unbalanced asymmetric information.

Insurance contracts made based on asymmetric information may attract people with high-risk type and drive people with low-risk type out of market. According to Akerlof (1970), the problem caused by adverse selection harms efficiency and leads to market collapse in some extreme case. Also, the existence of moral hazard makes the problem even worse. Moral hazard means that those who have insurance are less likely to take any preventative

activities and the probability of loss occurrence can be consequently increased or if loss occurs beneficiaries tend to over-use health care.

Medigap insurance has been accused by some policymakers since it encourages beneficiaries to consume more health care than what they actually need and therefore the expenditure of Medicare is higher than it should be. Especially some Medigap plans offer first dollar coverage, and beneficiaries who have incentive will tend to consume unnecessary services covered by Medicare. In order to reduce the growth rate of Medicare spending, some policymakers propose to discourage the purchase of Medigap coverage.

1.4 Research Questions

In this study, we seek to answer several questions by analyzing the MEPS (Medical Expenditure Panel Study) data in 2009, 2010 and 2011 respectively.

First, do individuals assess their risk in the correct way? We examine whether individuals have private information on their risk type to find the answer of this question.

Second, on average, who are more likely to purchase Medigap insurance--those with high risk type or low risk type? We study the statistical relationship between Medigap insurance purchase and self-rated health factors for this.

Third, is there any moral hazard in Medigap insurance market, does it have large impact if existed? We explore moral hazard indirectly by seeking to answer the previous two questions.

1.5 Organization of Dissertation

This dissertation is organized as follows. In Chapter 2, we review the previous literature that investigates selection effect and Medigap insurance market. The theoretical

model is elaborated in Chapter 3. In Chapter 4 and 5, we describe the data source, study sample selection and explain log-linear regression model, Probit model, seemingly unrelated regression model and EM algorithm. Data analysis and corresponding empirical results are shown in Chapter 6. Some potential problems are discussed in Chapter 7 and finally we conclude the dissertation in Chapter 8.

CHAPTER 2 LITERATURE REVIEW

2.1 Health Insurance Market

Wolfe and Goddeeris (1991) use Retirement History Survey (RHS) from 1977 to 1979 with 2,059 observations to investigate moral hazard and selection effect in Medigap insurance market. Three dimensional ex post loss, including hospital expenses, physician care expenditure and prescription drug expenses has been examined. They claim that evidence of adverse selection is found and the size of the effect is so small that it does not lead to severe efficiency problem. Also, they find that there is strong positive correlation between wealth and insurance demand.

Fang, Keane and Silverman (2008) investigate Medigap insurance market by combining data sets of Medicare Current Beneficiaries Survey (MCBS) and Health and Retirement Survey (HRS). They find that the total medical expenditure is $4,000 less for those with Medigap coverage comparing with those without Medigap coverage conditional on Medigap price. When controlling health status in addition to Medigap price, those with Medigap coverage tend to spend $2,000 more comparing with those without Medigap coverage. They claim that these two findings together indicate that there is advantageous selection in Medigap insurance market. Furthermore, they point out that the cognitive ability plays an important role in leading to the advantageous selection. The factor of risk preference, which has been focused and discussed for a long time, seems to be of less importance in Medigap market.

Buchmueller et al. (2006) examine private health insurance market in Australia. They find that individuals with private health insurance are typically in better health status

and have less physician visits comparing with those without private health insurance, and there is no significant difference of hospital utilization between those with and without private health insurance. Therefore, they claim that their findings support advantageous selection rather than adverse selection due to strict underwriting regulation. They show the evidence that risk aversion and income are the factors that lead to advantageous selection.

Ciro Avitabile (2009) uses data of 2004 and 2006 from Survey of Health, Ageing and Retirement in Europe to study the selection effect of individuals between age 50 and 75 in eight European countries. They use bivariate Probit model to analyze the data, and there is no significant positive correlation between the probability of owing private insurance coverage and the probability of ex post risk in terms of hospital treatment. However, they find that the survival probability is significantly positive correlated with the probability of signing a private hospital insurance and significantly negative correlated with the probability of overnight stay in hospital. This implies the existence of advantageous selection. Also, they claim that education and cognitive skills are two potential sources of advantageous selection.

Valetino Dardanoni and Paolo Li Donni (2012) study incentive and selection effect in Medigap insurance market using Health and Retirement Study (HRS) in 2002, 2004 and 2006. They examine the correlation between Medigap coverage and the utilization of health care in terms of inpatient care. They find that modest incentive effect and significant selection effect include both adverse selection and advantageous selection since residual heterogeneity is multidimensional.

2.2 Life Insurance Market

Correlation between life insurance and morality rate is analyzed in Mahdavi (2005). They consider both risk type and risk aversion and find that processing cost could affect individuals' choice of life insurance. They claim that if individuals are sufficient risk averse to reduce the mortality rate and processing cost is large enough, advantageous selection will show up in the insurance market, and if risk aversion is not sufficient enough to decrease the mortality rate to certain critical value and processing cost is not large enough, there would be adverse selection in the insurance market.

Yao, Schmit and Sydnor (2012) use data from the developing country Pakistan to analyze the relationship between claim history and renewal decision in a micro health insurance market. They find that households with more claim amount filed in the previous period are slightly more likely to renew their insurance in current period. However, comparing to those who are newly enrolled in current period renewed households are significantly less risky due to less claims filed and less amount of claim filed. Therefore, the risk type evolves with time and this scenario is advantageous in the insurance market.

Cutler and Zeckhauser (1997) empirically investigate importance of selection effect in two entities. One is Group Commission of Massachusetts, and the other is Harvard University. Harvard University provides equal contribution across plans. Under the equal contribution rules, PPO is driven into death spiral and crashes finally within three years due to adverse selection effect. Group Insurance Commission of Massachusetts subsidizes 85% premium regardless of plan cost, which does not result in death spiral of certain plan. However, it still leads to severe adverse selection effect.

2.3 Long-Term Care Insurance Market

Finkelstein and McGarry (2006) study the long-term care insurance market. Significant correlation between long-term care insurance and long-term care use is not found. However, they argue that lack of correlation does not necessarily mean that there is no asymmetric information. They show evidence that there is adverse selection on risk type and advantageous selection on risk preference. In general, there are two groups of people who are more likely to purchase long term care insurance: individuals who are high-risk type and individuals who are highly risk averse. The latter ones are less correlated with risk occurrence. Therefore, in this work, the risk almost cancels out with each other and it looks like there is no correlation. They also suggest a general framework to test asymmetric information in insurance market, which will be adopted in our dissertation as well.

2.4 Automobile Insurance Market

Chiappori and Salanié (2000) investigate asymmetric information in French auto insurance market. They concentrate on less experienced drivers in order to keep the individuals more homogeneous in terms of knowing their risk type. Using independent Probit model, bivariate Probit model and nonparametric model, they find no evidence to support existence of adverse selection.

Cohen and Einav (2005) study Israel auto insurance market. They examine both less experienced and experienced drivers. Their findings for less experienced drivers are consistent with those in Chiappori and Salanié (2000). It seems that less experience drivers in Israel do not have advantage on their risk type comparing with insurance company could

obtain. Therefore, there is no evidence for existence of adverse selection in less experienced drivers. However, it is another story for the experienced drivers. They do find that those with over three years driving experience have more private information than those insurance companies can control. They find that there is positive correlation between average number of claims and insurance coverage even after controlling many characteristics of policyholders.

CHAPTET 3 THEORTTICAL TERMINOLOGIES

3.1 Selection Effect

Rothschild and Stiglitz (1976) develop models that analyze market equilibrium under imperfect information through studying imperfect information in insurance market. They assume that the individual heterogeneity in terms of the probability of loss occurrence is alone. Charging the same averaged insured premium for different risk types would gradually drive those with relatively low risk out of the market and finally the whole market collapses. Therefore, pooling equilibrium does not exist in imperfect insurance market. If there were market equilibrium, it would be a separate equilibrium. In the separate equilibrium, those with high-risk type buy full coverage insurance and those with low-risk type buy partial coverage insurance or those with high-risk type buy insurance and those with low-risk type choose not to buy insurance. This separate market equilibrium demonstrates that levels of insurance coverage are adversely selected by individuals, and this is characterized as "positive correlation" between risk type and insurance coverage.

Later, this "positive correlation property" forms the basis for testing of selection effect by extensive empirical literature in different insurance markets. In some of them there is evidence to support this positive correlation, but there is not in others. Thus, the assumption that there is only uni-dimensional private information on risk type has been questioned by the seemingly contradicting results.

Hemenway (1990) modifies the theory by introducing the concept propitious selection, which emphasizes another piece of private information--the taste for risk. He summarizes the assumptions for adverse selection in this way: insurance purchase is not

mandatory; individuals are heterogeneous in risk type and this is known by themselves but not known by insurance companies. Therefore, insurance companies charge different risk-typed individuals with the same price. In other words, it means that insurance companies are not able to screen individuals, and propitious selection needs two more assumptions on top of the assumptions for adverse selection. One is that individuals are heterogeneous in tastes for risk and their taste are consistent across physical and financial dimensions. The other is that individuals could affect their risk by positive or negative actions or attitude in some extent.

De Meza and Webb (2001) theoretically show a similar idea, and they define advantageous selection in basis of individuals heterogeneous in both risk and risk preference. Also, risk preference is endogenous with risk precautionary activities.

In general, any private information that is positive correlated with insurance coverage and negative correlated with risk could potentially lead to negative correlation between risk type and insurance coverage if its effect is large enough.

Therefore, in order to investigate selection effect, it is paramount to know whether private information is one-dimensional or multi-dimensional. When there is multi-dimensional private information, for instance, besides risk type there is risk preference (not limited to risk preference) that is positively correlated with insurance coverage and negatively correlated with risk type, this could attenuate the positive correlation between risk type and insurance coverage. If it is large enough, it could even reverse the sign of the correlation between risk type and insurance coverage. Suppose there were no moral hazard in this case, then average expenditure of individuals with insurance is lower than the

average expenditure of those without insurance , which is defined as advantageous selection (Fang et al, 2008)

Suppose there is pure selection effect and risk type is the only dimensional private information of individuals, individual's choice of insurance coverage can be described by the following formula,

$$C = f(r|I),$$

where C is level of insurance coverage, I is information that individuals have and r stands for risk types. The function of f is monotonically increasing function with respect to r conditional on I.

Individual evaluates his/her risk type and makes decision on choosing insurance coverage accordingly. When information is more than one-dimensional, the corresponding formula can be described in this way:

$$C = g(r, \boldsymbol{x}|I),$$

where $\boldsymbol{x}$ is a vector that includes all characteristics other than risk type, which correlates with choice of insurance coverage and risk type. Therefore, the insurance coverage level is determined not only by risk type but also other factors when certain individual makes choices of insurance.

This is more complicated comparing with the one-dimensional private information case, and relying on positive correlation test would be not enough to handle the situation with more than one-dimensional private information since risk type increase does not necessarily increase insurance coverage level. Positive correlation test between risk type

and insurance coverage under this circumstance without considering other potential factors will be also problematic due to endogeneity problem.

Empirically, individuals' risk type is a latent variable. Testing correlation between risk type and insurance coverage exists only in ideal theoretical case. In realistic cases, frequently used variable of proxy risk type is realized loss. The issue introduced by using realized loss is that realized loss reflects not only selection effect but also moral hazard if there is any. Not rejecting existence of positive correlation between insurance coverage and realized loss could not tell us that the correlation is derived from pure adverse selection, pure moral hazard or both without further investigation.

3.2 Moral Hazard

Moral hazard can be divided into ex ante moral hazard and ex post moral hazard. Generally speaking, it reflects an incentive effect due to having insurance. In the case of ex ante moral hazard, health insurance companies would give policyholders incentive to take less precautious activities before occurrence of loss, and if policyholders use more health care than what they actually need after loss occurrence, ex post moral hazard occurs.

Ex ante moral hazard endogenously increases risk type. Ex post moral hazard leads to over consumption of health care. However, it does not essentially change the risk type. Ex ante moral hazard appears more often in market like automobile insurance market and less often in health related insurance market. Therefore, in this empirical work we assume there is no ex ante moral hazard and whenever moral hazard is mentioned, we mean ex post moral hazard.

3.3 Selection Effect and Moral Hazard

The whole story of moral hazard and selection effect can be described as follows. First of all, individuals select insurance coverage according to their information of risk type, risk preference and/or other relevant factors they have. Some of them choose full coverage, some of them choose less coverage, and others may end up no coverage. After choices being made, if those who have insurance become more reckless (choosing less healthy) comparing those who have no insurance increases their risk type and this situation is ex ante moral hazard. Once loss occurs, insured individuals consume more health care than what they actually need due to holding the coverage of health insurance and ex post moral hazard happens. The timeline of the whole procedure is described in Figure 3.1.

If adverse selection exists in the insurance market, the existence of moral hazard will strengthen the positive correlation between insurance coverage and ex post loss. If there is advantageous selection in the insurance market, the existence of moral hazard will attenuate its effect. Positive or negative overall correlation depends on whether the magnitude of advantageous selection is larger than the moral hazard or not.

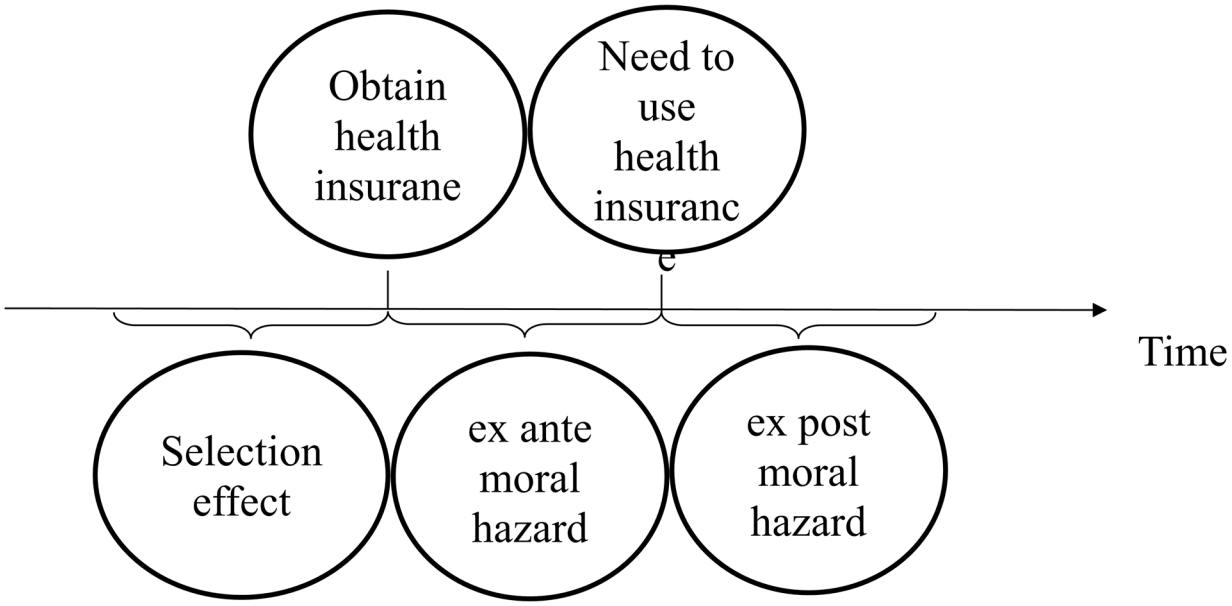


Figure 3.1 The Timeline of Selection Effect and Moral Hazard

CHAPTER 4 DATA

4.1 Medical Expenditure Panel Survey

Medical Expenditure Panel Survey (MEPS) is sponsored by the Agency for Healthcare Research and Quality started in 1996. There are two major components in MEPS. One is called Household Component (MEPS-HC or HC), and the other is called Insurance Component (MEPS-IC or IC). MEPS-HC is a subsample of households that participated previous year's NHIS (National Health Interview Survey). Each year a new panel of households is selected, and it is followed for two consecutive years with 5 rounds of interviews of non-institutionalized U.S population. We use MEPS-HC data in our study. The data includes demographic information, health care services, health status, health care expenditures, insurance information, employment status, income, etc.

4.2 Sample selection and Sample size

In this study, we compare respondents, who are older than 65, have Medicare Part A and Part B but do not have Medigap insurance with respondents, who are older than 65, have Medicare Part A and Part B and choose to purchase Medigap coverage voluntarily.

Initially, we identify those who are covered by Medicare Part A and Part B, since Medigap is the supplemental insurance for Medicare. Only individuals that are covered by Part A and Part B are eligible for purchasing Medigap coverage. Then individuals, who have Medicare Part A and Part B and do not have any private insurance, are considered as no Medigap coverage (coded as Medigap=0). Individuals with Part A and Part B, who report to have Medigap and pay the premium fully on their own or by family members, are considered as having Medigap (coded as Medigap=1). We exclude those who have

Medigap coverage and whose part of or the entire coverage premium is paid by their/their spouse's employers or their/their spouse's former employers, since it is ambiguous whether individuals would purchase Medigap coverage without subsidy from their/their spouse's employers or their/their spouse's former employers. Detailed conditions for construction of variable of Medigap are shown in Table 4.1.

MEPS collect a new panel of sample households each year, and this panel of households are followed for two calendar years. It is an overlapping panel design. Therefore, for each calendar year you can observe two different groups of sample. In this study we will not take advantage of the panel design. We use data from the year of 2009, 2010 and 2011 and only do cross-sectional analysis without stacking different years' data together.

4.3 Variables

4.3.1 Dependent Variables

One dependent variable we use in the regression model is the total health expenditure. The range of the total health expenditure is [0, 247,828] in 2009, [0, 207,213] in 2010 and [0, 209,265] in 2011 respectively. All the data points in the sample are non-negative and sample distribution is skewed. Therefore, in order to make the distribution of the data better to fit the normal distribution assumption, we take the logarithm of the total health expenditure later in the regression analysis.

There are 113 observations with total expenditure equal to 0 in 2009, 122 in 2010, and 126 in 2011 respectively. Taking the logarithm of them will make the value negative infinity. In order to satisfy the normal distribution assumption of our linear regression

model and remove the logarithm singularity of zero, we smooth the point mass distribution of observations with zero total expenditure by assigning the values randomly drew from a uniform distribution ranged from zero to the nonzero minimum value. The sample distribution from year 2009 before and after taking logarithm with our above modification is shown in Figure 4.1.

Another dependent variable is purchasing or not purchasing Medigap coverage. It is a binary variable with 1 reflecting holding Medigap coverage and 0 reflecting not holding Medigap coverage.

4.3.2 Independent Variables

Age, gender and *smoking status* are the variables that used by insurance company to set the premium. Thus these three variables and their interactions are added in the model in order to control the risk classification evaluated by insurance companies. We call this group of variables as "premium variables" from now on since they are used to set the premium by the insurance companies for Medigap coverage.

Self-rate health: this is a subjective health rate. It shows how would individuals rate their health status, and it is an ordinal categorical variable.

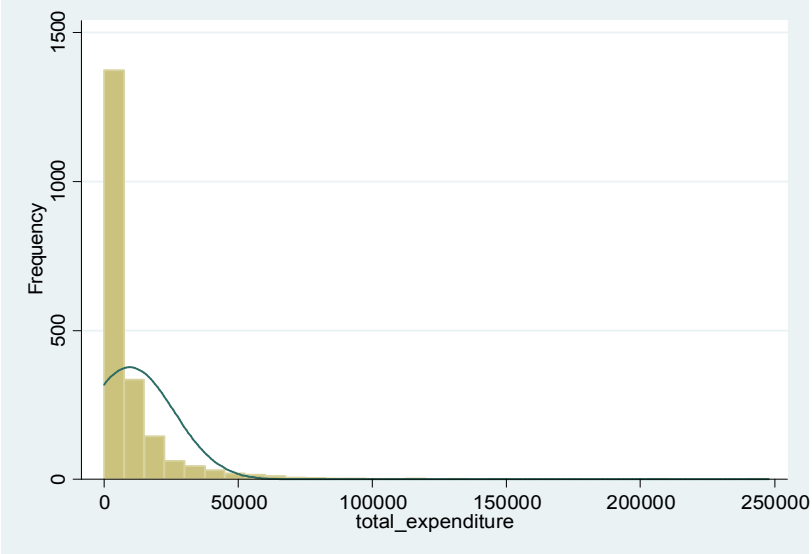
Poverty categories: it reflects the family income as percentage of poverty line. It directly shows the economic status of individuals.

Education: it shows the years of education when initially enrolling in MEPS.

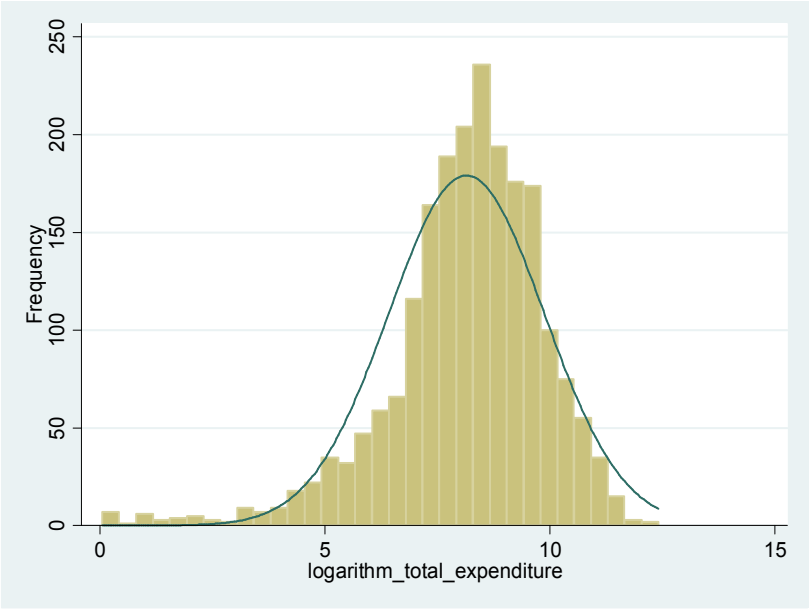
Cognitive limitations: it shows whether individuals have limitations in cognitive ability.

Risk Aversion: it measures degree of individuals' risk aversion.

Risk aversion, cognitive limitations, education and poverty categories are denoted as "additional variables" for short in later chapters. All descriptions of variables are shown in Table 4.2.



A. Sample Distribution before Taking Logarithm



B. Sample Distribution after Taking Logarithm

Figure 4.1 Sample Distributions of the Total Health Expenditure and the Logarithm of the Total Health Expenditure of 2009 data

Table 4.1 Construction of Variable Medigap

Are you currently covered by Medicare health insurance?	Yes	Yes
Part A of Medicare covers most hospital expenses. Part B covers many doctors expenses including doctor visits, and the premium is usually deducted from your Social Security. Are you covered under Part B of Medicare?	Yes	Yes
Do you have Medigap Insurance?	Yes	No
Including any help from your family, do you (or your (husband/wife/partner) pay all of the costs, some of the costs, or none of the costs of the premium for this health insurance coverage?	All	N/A
Binary Variable "Medigap"	1	0

Table 4.2 Descriptions of Variables

Name of Variables	Variables Definition
Age	Age in Years
Female	1 if female, 0 otherwise
Smoke	Currently smoke, 1 if Yes, 0 Otherwise.
Education	Years of Education 1 - 8 ELEMENTARY GRADES 1 - 8 9 - 11 HIGH SCHOOL GRADES 9 - 11 12 GRADE 12 13 1 YEAR COLLEGE 14 2 YEARS COLLEGE 15 3 YEARS COLLEGE 16 4 YEARS COLLEGE 17 5+ YEARS COLLEGE
Married	Marital Status, 1 if married, 0 Otherwise.
Race	White, Black or Others
Expenditure	Total health expenditure
Risk Aversion	More likely to take risks 1 Strongly disagree 2 Somewhat disagree 3 Uncertain 4 Somewhat agree 5 Strongly agree
SR Health	Self-rated health status 1 Excellent, 2 Very Good 3 Good 4 Fair 5 Poor
Medigap	Hold Medigap coverage, 1 if Yes, 0 Otherwise
PW	Person Weight
Cognition	Cognitive limitations, 1 if Yes, 0 Otherwise.

4.4 Survey Data

4.4.1 Sampling

Simple random sampling is based on the assumption that each individual in the population has same probability to be selected into the sample. Survey data is rarely collected as simple random sample due to the cost or time saving reason. MEPS data is also not an exception. The data collected in MEPS-HC is based on a complex survey design. Therefore, we cannot treat data with complex survey design as simple random sample to reflect the general population. Weight, cluster and strata should be handled in the analysis to approximate a representative sample. Misspecification of the sampling design could possibly result in incorrect point estimates with bad standard errors.

4.4.2 Sample Weight, Cluster and Stratification

Sample weight is the inverse of the likelihood of being sampled when sampling is not random. When collecting data, some group may be under-sampled due to no response and some individuals and minority group may be over-sampled on purpose. Therefore, when sampling statistics are calculated, weighting adjustment is required. For example, the total initial respondents in our study are 2,344 for 2009, 2,274 for 2010, and 2,314 for 2011 respectively. Due to the 0 sample weight, sample size is adjusted to 2,295 for 2009, 2,213 for 2010, 2,256 for 2011 respectively, which is shown in Table 4.3.

In addition, when MEPS data are collected, cluster and stratum are used to save time and improve efficiency. In cluster sampling, a group of elements instead of a single element of population is used as primary sample units. Smaller group of elements or single elements are called secondary sample units, which depends on it is multi-stage sampling or

single stage sampling. In stratified design, population is sampled into well-defined group, and elements are independently sampled from each stratum. Cluster sampling usually lead to larger variance compared to simple random sampling and stratification usually lead to smaller variance compared to simple random sampling.

Both primary sample units (PSU) and stratification are used in MEPS data collection so that in our analysis we need to adjust variances and standard errors correspondingly.

Table 4.3 Detailed Data Sample Information

	2009	2010	2011
Sample Size	2,344	2,274	2,314
Sample Size Used	2,295	2,213	2,256
Sample Sized Skipped Due To Non-positive Weight	49	61	58
Number of Strata	164	164	162
Number of PSU	327	326	322
Weighted Size for Used Observations	24,070,044	24,662,204	24,729,641

CHAPTER 5 STATISTICAL MODELING

5.1 Conventional Positive Correlation Test

As we have mentioned in Chapter 3, the positive correlation test investigate reduced form correlation between risk type and insurance coverage. When the private information is more than one dimensional, the result from model without controlling other informative factors can be misleading. However, positive correlation test still can be a good starting point.

5.1.1 Multiple Linear Regression Analysis

Our regression model of positive correlation test regressing logarithm of the total expenditure on Medigap coverage and premium variables is as follows:

$$Y_{TE} = \alpha Y_M + X_{premium}\beta + \varepsilon \qquad (1)$$

where Y_{TE} is the logarithm of the total health expenditure for each year, Y_M is the binary variable reflects holding or not holding Medigap coverage and $X_{premium}$ is the vector of premium variables, including age, gender and smoking status and their interactions. This is the reduced form relation between Medigap coverage and the logarithm of the total health expenditure. The coefficient of α reflects the correlation between the logarithm of the total health expenditure and Medigap coverage when controlling premium variables for risk classification.

We use this model to test the following hypothesis:

Hypothesis I: there exists the nonzero positive correlation between Medigap coverage and total health expenditure when controlling premium variables.

In the above hypothesis, the null side H_o is zero correlation and the alternative one H_1 is nonzero positive correlation.

Table 5.1 Coefficients and Standard Errors of variable Medigap from Positive Correlation Test in Year 2009, 2010 and 2011

	2009	**2010**	**2011**
Medigap	0.3855*** (0.1228)	0.1642 (0.1842)	0.2014 (0.2535)

1) Standard error in parentheses

2) Asterisks indicate level of statistical significance: ***p<0.01, **p<0.05, *p<0.1

As shown in Table 5.1, we find a positive correlation between the logarithm of the total expenditure and Medigap in all years of 2009, 2010 and 2011. According to the statistical p-values, we cannot reject the null hypothesis for 2010 and 2011, but we are able to reject null hypothesis at 1% significant level for 2009.

Without further investigations, there are several possible ways to explain the coefficients in the regression outcomes from the data of 2009, 2010 and 2011. Firstly, there could be no private information or individuals could decide not to use this only dimensional private information to guide their choice of Medigap coverage. Therefore, the positive correlation result only reflects moral hazard. For 2010 and 2011, the magnitude of moral hazard is small and not significant at any conventional significance level. For 2009, the coefficient of Medigap is 0.3855 and significant at 1% level (Table 5.1), which means total health expenditure of those holding Medigap coverage is 47% higher than those without Medigap due to incentive effect. Secondly, another possibility is that individuals do have private information about their risk type and are able to use it, but the magnitude of that information is mild, and only significant for 2009, and there is no moral hazard. Thirdly, one more possibility is that there exist both adverse selection and moral hazard, and both of them are really small. Finally, the last possibility is that according to multi-dimensional private information, private information includes but is not limited to individuals who know their risk type, and those individuals have private information in other dimensions. This results in advantageous selection. Advantageous selection offsets the effect or at least partial effect of moral hazard.

We have mentioned early that Medigap insurance market is a highly regulated by the government. Highly regulation on underwriting is usually proved to lead to severe asymmetric information. Therefore, the assumption that there is multi-dimensional private information sounds more reasonable. Of course, we need further empirical data investigation to prove our thoughts.

Also, if there were indeed factors rather than risk type, which are correlated with Medigap and the total health expenditure, relying on positive correlation test would cause endogeneity problem when including the related factors in the error term.

5.2 Independent Probit model and multiple regression model

Fortunately, we have a variable called self-rated health in MEPS. It describes how individuals rate their health status. It is a natural proxy variable for self-rated risk type. Therefore, rather than using positive correlation test we follow the basic idea of Finkelstein and McGarry (2006). On one hand, we estimate the relationship between self-rated health and his/her subsequent health expenditure conditional on risk classification (in other words, conditional on premium variables) to see whether individuals correctly assess their risk level. On the other hand, we examine relationship between self-rated health and his/her Medigap insurance holdings.

Baseline Specification

We estimate the following two equations independently.

$$Y_{TE} = X_{srh}\alpha + X_{premium}\beta + \varepsilon_{TE} \qquad (2)$$

$$Y_M^* = X_{srh}\alpha + X_{premium}\beta + \varepsilon_M \qquad (3)$$

where Y_M^* is the latent variable for the binary variable Medigap Y_M in the Probit model, i.e., $Y_M = I(Y_M^* > 0)$. The corresponding two hypotheses we would like to test are:

Hypothesis II: The regression coefficient of self-rated health status is nonzero.

The null side of the above hypothesis is zero coefficient of self-rated health status, and the alternative one is nonzero coefficient of self-rated health status.

Hypothesis III: The correlation between Medigap coverage and self-rated health is nonzero.

In this hypothesis, we have the null side with the zero correlation between Medigap coverage and self-rated health is zero, and the alternative one is the nonzero correlation.

Self-rated health status is an ordinal categorical variable. In the Probit model, the group of individuals who report that they are "excellent" in health status is omitted. Comparing with those who report "excellent", individuals who report "very good" are less likely to have Medigap coverage. After controlling premium variables if risk type is the only private information left for individuals to make decisions then we expect less risky individuals choose less health coverage. If not, it implies the private information is not uni-dimensional. In that case we would like to see what other factors that could affect choice of coverage and the logarithm of the total expenditure. The extensive specification is as follows.

Extensive Specification

$$Y_{TE} = X_{srh}\alpha + X_{premium}\beta + X_{additional}\gamma + \varepsilon_{TE} \quad (4)$$

$$Y_M^* = X_{srh}\alpha + X_{premium}\beta + X_{additional}\gamma + \varepsilon_M \quad (5)$$

5.3 Seemingly Unrelated Regressions and EM algorithm

In the independent models, one issue that has not been addressed is there could be correlation between error terms in the two equations. If the error terms are not correlated, then estimate the two equations jointly or independently would equally efficient. If the error terms of two equations are correlated, jointly estimating the two equations simultaneously is more efficient. Seemingly Unrelated Regressions (SUR) are proposed for this purpose and it is based on the assumption that two dependent variables are continuous variables or at least can be treated as continuous variables. Bivariate Probit model is proposed to handle the situation when the two jointly estimated equations are with binary dependent variables. In our model, Medigap coverage is a binary variable and the logarithm of the total health expenditure is continuous; therefore, neither SUR nor bivariate Probit model is ready to be used without any adjustment in our case.

In order to fit one continuous variable and one binary variable in SUR model, we first estimate the continuous latent variable in the Probit model. Then we use estimated latent variable and the logarithm of the total expenditure as the two continuous dependent variables in the SUR. We adopt EM algorithm to estimate the latent variable.

EM (expectation-maximization) algorithm is one iterative statistical method to estimate parameters of statistical models with maximizing the expected value of log-likelihood. It has wide application to estimate parameters with missing variables (Dempster, Laird. and Rubin, 1977). In brief, EM-iterations include two steps: an expectation (E) step and a maximization (M) step. Suppose that X is the observed variables and Z is the unobserved hidden variables, and we need to estimate the parameter θ in the

statistical model with likelihood function $L(\theta; X, Z)$. In the E-step, we calculate the expected value of the log-likelihood function Q with respect to the conditional distribution of Z given X at the t^{th} step with $\theta^{(t)}$:

$$Q\left(\theta\middle|X, \theta^{(t)}\right) = E\left[logL(\theta; X, Z)\middle|X, \theta^{(t)}\right].$$

Then in the M-step, the maximization of Q is calculated to find the next step $\theta^{(t+1)}$:

$$\theta^{(t+1)} = argmax_{\theta} Q\left(\theta\middle|X, \theta^{(t)}\right).$$

For Probit model, we model the observed Y as $I(Y^* > 0)$ where Y^* is the hidden variable and furthermore we can assume that $Y^* = X\beta + \varepsilon$, where $\varepsilon \sim N(0,1)$. Since we only observe Y and Y^* is hidden, we can use EM algorithm to estimate the parameter β in the Probit model. In the E-step, we have

$$Q\left(\beta\middle|X, Y, \beta^{(t)}\right) = E\left[logL(\beta; Y, Y^*)\middle|X, Y, \beta^{(t)}\right]$$

$$= -\frac{n}{2}\log(2\pi) - \frac{1}{2}E[\| Y^* - X\beta \|^2 |X, Y, \beta^{(t)}]$$

$$= -\frac{n}{2}\log(2\pi) - \frac{1}{2}E\left(Y^{*\prime}Y^*\middle|Y, \beta^{(t)}\right) + \beta' X' E\left(Y^*\middle|Y, \beta^{(t)}\right) - \frac{1}{2}\beta' X' X\beta$$

Denote that $\mu\left(Y, \beta^{(t)}\right) = E\left(Y^*\middle|Y, \beta^{(t)}\right)$. In the M-step, we maximize $Q(\beta|X, Y, \beta^{(t)})$ and have

$$\beta^{(t+1)} = (X'X)^{-1}X'\mu\left(Y, \beta^{(t)}\right),$$

which is the exact linear regression solution after regressing $\mu(Y, \beta^{(t)})$ on X. When $Y = 1$, we have

$$\mu\left(Y = 1, \beta^{(t)}\right) = E\left(Y^*\middle|Y = 1, \beta^{(t)}\right)$$

$$= E(X\beta + \varepsilon|Y = 1, \beta^{(t)})$$

$$= X\beta^{(t)} + E(\varepsilon|\varepsilon > -X\beta, \beta^{(t)})$$

$$= X\beta^{(t)} - E(\varepsilon|\varepsilon < X\beta, \beta^{(t)})$$

$$= X\beta^{(t)} - \frac{\int_{-\infty}^{X\beta^{(t)}} \frac{z}{\sqrt{2\pi}} e^{-z^2/2} dz}{\Phi(X\beta^{(t)})}$$

$$= X\beta^{(t)} + \frac{\int_{-\infty}^{X\beta^{(t)}} \frac{1}{\sqrt{2\pi}} de^{-z^2/2}}{\Phi(X\beta^{(t)})}$$

$$= X\beta^{(t)} + \frac{\phi(X\beta^{(t)})}{\Phi(X\beta^{(t)})},$$

where $\phi(z)$ and $\Phi(z)$ are the probability density function (PDF) and the cumulative density function (CDF) of the standard normal distribution respectively. Similarly we have

$$\mu\big(Y = 0, \beta^{(t)}\big) = X\beta^{(t)} - \frac{\phi\big(X\beta^{(t)}\big)}{1 - \Phi(X\beta^{(t)})}.$$

Therefore, in general, we have

$$\mu\big(Y, \beta^{(t)}\big) = X\beta^{(t)} + \frac{\phi(X\beta^{(t)})}{\Phi(X\beta^{(t)})[1 - \Phi(X\beta^{(t)})]}\big[Y - \Phi\big(X\beta^{(t)}\big)\big].$$

In our statistical model, denote the logarithm of the total expenditure observation as Y_{TE} and Medigap observation as Y_M. We need to consider the correlation structure between two variables. We model Y_M through the Probit model, the correlation now can be solved through seemingly unrelated regression (SUR) between Y_{TE} and Y_M^*:

$$Y_{TE} = X\beta_{TE} + \varepsilon_{TE}, \quad (6)$$

$$Y_M^* = X\beta_M + \varepsilon_M, \quad (7)$$

where the correlation $\sigma_{TE,M}$ between ε_{TE} and ε_M is estimated. By considering SUR and the Probit model jointly, we can revise the EM algorithm for the Probit model to incorporate SUR as the following:

- E-step: find the expected Y_M^* with respect to X, Y and $\beta_M^{(t)}$:

$$\mu\left(Y_M, \beta_M^{(t)}\right) = X\beta_M^{(t)} + \frac{\phi(X\beta_M^{(t)})}{\Phi\left(X\beta_M^{(t)}\right)[1-\Phi\left(X\beta_M^{(t)}\right)]}\left[Y - \Phi\left(X\beta_M^{(t)}\right)\right].$$

- M-step: use SUR with response variables Y_{TE} and $\mu(Y_M, \beta_M^{(t)})$ and predictors X to find $\beta_M^{(t+1)}$ and $\beta_{TE}^{(t+1)}$.

We run the above EM algorithm till converge when $\| \beta_M^{(t+1)} - \beta_M^{(t)} \| + \| \beta_{TE}^{(t+1)} - \beta_{TE}^{(t)} \| \leq 10^{-9}$.

In the following joint regression model,

$$Y_{TE} = X\beta_{TE} + \varepsilon_{TE},$$

$$Y_M^* = X\beta_M + \varepsilon_M,$$

we estimate the two regression model simultaneously by controlling premium variables only, and controlling both premium variables and additional variables. The correlation ρ between the two regression errors ε_{TE} and ε_M reflects relation between Medigap coverage and the total health expenditure conditional on corresponding controlling variables.

Note: Equations (6) and (7) are simplified version of Equations (2) and (3) and Equations (4) and (5) are for the sake of formula derivation.

CHAPTER 6 ANALYSIS

6.1 Independent Multiple Linear Regression Model and Probit Model

Self-rated health

In Table 6.4, by conditioning on only premium variables (*sex, age* and *smoking status*), we can see that there is significant increase of the logarithm of the total health expenditure when self-rated health changes from "excellent" to "poor". Comparing to "excellent", the total expenditure of those with "very good" self-rated health is around 36.86% higher on average, the total expenditure of those with "good" self-rated health is around 65.76% higher, the total expenditure of those with "fair" is around 107.3% higher and the total expenditure of those with "poor" self-rated health is almost 189% higher.

Controlling additional variables such as risk aversion, cognitive limitations, education and poverty categories, we still find strong and significant pattern of self-selected health status on the logarithm of total health expenditure.

It demonstrates that individuals are in general quite informed about their health status conditional risk classification assessed by insurance companies and there is still large residual private information. Meanwhile, much stronger positive correlation shows after controlling premium variables plus additional variables. It implies that this is private information exists across individuals from different social status level and this is even much stronger within each group. Similar pattern holds for the results of both 2010 data and 2011 data as we can see from Table 6.5 and Table 6.6.

In Table 6.7, using Probit model with Medigap as dependent variable, we do not find strong correlation between self-rated health with Medigap purchase. All of the

coefficients are negative, however, most of them are not significant at any conventional significant level irrespectively of controlling only premium variables or premium variables plus additional variables. Only the coefficient of self-rated health at "Very Good" is significant at 10% level.

From Table 6.8, Table 6.9 of both 2010 data and 2011 data, we find the coefficients are negative and most of them are significant when controlling premium variables or premium variables plus additional variables.

According to the information provided from Table 6.4 to 6.9, we find that individuals are quite informed about their health status and they have private information that insurance companies do not use. At the same time self-rated health is not positive correlated with Medigap purchase regardless of controlling only premium variables or controlling premium variables, risk aversion, cognitive limitation, education, and poverty categories as well. In other words, we find individuals have private information of their risk type, but fail to find the existence of adverse selection conditional on these variables.

Next, we will examine the effect of several potential factors that have been frequently appeared in study of insurance market selection effect.

Risk Aversion

Risk aversion is a factor, which is most frequently examined in the previous studies in various insurance markets. It shows that risk aversion plays an important role in individuals' choice of insurance in some cases. Those who are very risk averse are more likely to purchase insurance even though their risk type is very low.

In MEPS, individuals are asked to rate their "likelihood to take risk". Answers of risk aversion are leveled from "disagree strongly", "disagree somewhat ", "uncertain" "agree somewhat" to "agree strongly". As we can see from the Table 6.4 to Table 6.6 most of the coefficients are negative and at least half of them are not significant at any conventional level. Therefore, it implies that those individuals who are more risk averse do not necessarily have less health expenditure than those with less risk averse.

There could be several possibilities for this scenario. In one possible situation, individuals with higher tolerance of risk are sick, but there are less likely to see a doctor and then less health expenditure occurs. If that is the case, their lower expenditure does not reflect that they are lower in risk. Another possibility is that risk aversion is not universal across various aspects. Willing or not willing to take risk in general does not consistently reflect individuals' attitude toward risk of health. This could also explain why we do not see that individuals with risk averse in general case do not have less health expenditure.

In addition, from the results of the Probit models we can find that more risk aversion does not seem to systematically correlate with Medigap purchase either. In 2009 data, the coefficients of risk aversion at different level are negative. In 2010 data, most of the coefficients of risk aversion are positive. All of the coefficients are not significant at any significant level. Either risk aversion is not that important in making choice of Medigap insurance or risk aversion surveyed in MEPS are not reflecting individuals attitude toward health risk.

Cognitive limitation and Education

Individuals with cognitive limitation have significantly higher total health expenditure than those without cognitive limitation. The percentage of increment ranges around 38.7% to 75.2%within these three years.

Effect of education is relatively small, which ranges from 5% to 11%. This means additional years' education ceteris paribus is positively correlated with total health expenditure. Generally speaking, we expect that better education would be correlated with better health status. Especially there are evidences about this in several previous studies. Possible explanation for this counter-intuition result could be that those with better education understand health and health insurance better than those with lower education level; therefore, ceteris paribus, those with better education tend to be more likely to utilize health service and therefore generate more health expenditure.

For Medigap purchase, we do not get consistent result from the data of 2009, 2010 and 2011 regarding cognitive limitations. In the data of 2009 and 2010, those with cognitive limitations are less likely than those without cognitive limitations to purchase Medigap insurance before we control poverty categories, and this is consistent with results of study from Fang, Keane and Silverman (2008). After we control poverty categories, the coefficients become not significant. We do not find such correlation between cognitive limitation and Medigap purchase in 2011 data.

Similar for education, in the data of 2009 and 2010, it shows that years of education and Medigap purchase are positively correlated and significant, and this is as expected. In the data of 2011, we do not find such correlation.

Poverty Categories

Economic status is also an important factor to look at in studying selection effect. Economics status is often found in some of former studies to have negative correlation with health expenditure and positive correlation with choice of health insurance. We do not find any significant correlation between total health expenditure and poverty categories in any year from 2009 to 2011.

In the data of 2009, we do not find significant correlation between the income level and Medigap purchase. In the data of 2010 and 2011, individuals with higher income level are significantly more likely than those with lower income level to purchase Medigap. For example, in the data of 2011, comparing to those who are in poor or negative category, the coefficient of those who are in high category is 0.8735 and significant at 1% level. It implies that those in high-income category have average marginal probability 8.5% higher than those in poor/negative income category. Other than that when we add poverty categories, coefficients of cognitive limitation become not significant at any conventional level in the data of 2009 and 2010.

6.2 Seemingly Unrelated Regression (SUR) Model

To examine whether there is correlation between errors between Medigap coverage and the total health expenditure, SUR model is more efficient.

Self-Rated Health

We still find significant positive correlation between the self-rated health status and the logarithm of total health expenditure, which implies individuals' private information of their risk type. Most of the coefficients of self-rated health on Medigap purchase in SUR become negative and significant. The magnitude of the coefficients from SUR and

independent models are very tiny. The major difference of using SUR and independent models are the shrinking of standard errors especially for self-rated health status on Medigap coverage.

Combining findings above, the scenario can be explained in this way. First, there is advantageous selection effect in Medigap insurance market. Since we first do not control the correlation between errors, we cannot reject that coefficient of self-rated health on Medigap purchase is different from 0 in independent model. After we do that, we immediately see the coefficients becomes significant.

Second, there is moral hazard in Medigap insurance market. Before we use SUR to control error terms in the logarithm of total expenditure and Medigap purchase, we do not find evidence of adverse selection in the data of 2009. After we control error term correlation in SUR, we find advantageous selection in all the three years. Existence of moral hazard mitigates the negative correlation between the total health expenditure and Medigap purchase due to advantageous selection.

Risk aversion

In the SUR models, we still do not observe consistent and stable effects from risk aversion.

Cognitive limitation and Education

Cognitive limitation is positively and significantly correlated with the total health expenditure in all three years, and it is negatively and significantly correlated with Medigap purchase only in the data of 2009 and 2010. In the data of 2011, the correlation between cognitive limitation and Medigap is almost zero and not significant. Therefore, if

cognitive limitation serves as the source of advantageous selection, it only works for 2009 and 2010.

Education is positively correlated with the total health expenditure and positively correlated with Medigap purchase in all three year. Thus, in this study we do not find years of education as source of advantageous selection.

Poverty Categories

Poverty categories do not have positive or negative correlation with health in terms of total the health expenditure. However, better economic status is associated with higher possibility to purchase Medigap. Thus, economics status does not work as source of advantageous selection as well.

Table 6.1 Descriptive Statistics-2009

	All	Without Medigap	With Medigap(premium paid on their own)	With Medigap (premium not paid on their own)
Age	74.55	74.47	74.87	74.92
Female	0.564	0.562	0.583	0.570
Smoke	0.091	0.099	0.065	0.054
Income	52,246.40	51,117.84	57,741.45	56,991.16
Education	12.240	12.043	13.379	12.967
Married	0.544	0.539	0.539	0.584
Race				
White	0.872	0.861	0.942	0.909
Black	0.080	0.086	0.040	0.059
Other	0.048	0.053	0.018	0.032
Races				
Expenditure	9,581.02	9,438.13	10,260.24	10,191.36
ER	0.238	0.228	0.233	0.312
Seatbelt	1.332	1.355	1.217	1.237
Risk Aversion	2.350	2.369	2.370	2.206
SR health	2.725	2.755	2.635	2.562
Cognition	0.106	0.114	0.068	0.069
Medigap		0	1	2
Sample size	2,295	1,945	128	222
Weighted size	24,070,045	19,652,559	1,624,905	2,792,580

Table 6.2 Descriptive Statistics-2010

	All	Without Medigap	With Medigap (premium paid on their own)	With Medigap (premium not paid on their own)
Age	74.595	74.577	75.384	74.398
Female	0.568	0.561	0.546	0.549
Smoke	0.097	0.099	0.083	0.087
Income	54,166.73	52,769.70	70,418.46	54,864.11
Education	12.448	12.276	13.238	13.298
Married	0.566	0.561	0.644	0.561
Race				
White	0.870	0.858	0.946	0.912
Black	0.082	0.090	0.023	0.058
Other	0.048	0.052	0.030	0.030
Races				
Expenditure	9,912.59	9,701.67	8,737.73	12,307.70
ER	0.246	0.245	0.209	0.279
Seatbelt	1.349	1.356	1.163	1.418
Risk Aversion	2.228	2.243	2.115	2.180
SR health	2.693	2.734	2.504	2.489
Cognition	0.097	0.105	0.054	0.066
Medigap		0	1	2
Sample size	2,213	1,914	118	181
Weighted size	24,662,204	20,415,159	1,643,182	2,603,863

Table 6.3 Descriptive Statistics-2011

	All	Without Medigap	With Medigap(premium paid on their own)	With Medigap (premium not paid on their own)
Age	74.216	74.298	74.360	73.632
Female	.562	.553	.579	.609
Smoke	.097	.102	.116	.055
Income	56,276.10	54,876.32	55,847.18	65,186.21
Education	12.453	12.230	13.255	13.548
Married	0.553	0.548	0.563	0.575
Race				
White	0.870	0.849	0.971	0.947
Black	0.086	0.099	0.013	0.042
Other	0.045	0.052	0.016	0.011
Races				
Expenditure	9,718.61	9,752.35	10,497.76	9,106.77
ER	0.232	0.231	0.196	0.257
Seatbelt	1.385	1.393	1.567	1.245
Risk Aversion	2.295	2.280	2.463	2.296
SR health	2.675	2.752	2.248	2.420
Cognition	0.101	0.110	0.069	0.062
Medigap		0	1	2
Sample size	2,257	1,921	109	227
Weighted size	24,729,641	19,874,074	1,653,731	3,201,835

Table 6.4 Log-linear Regression Results of 2009 Data

Log-linear regression with logarithm of total expenditure as the dependent variable

	Reg1	Reg2	Reg3	Reg4	Reg5
Self-rated health					
Excellent					
Very good	.3686***	.3698***	.3750***	.3911***	.3922***
	(.1287)	(.1267)	(.1272)	(.1272)	(.1266)
Good	.6576***	.6627***	.6579***	.7799***	.7882***
	(.1430)	(.1421)	(.1405)	(.1400)	(.1387)
Fair	1.073***	1.080***	1.024***	1.192***	1.198***
	(.1584)	(.1564)	(.1567)	(.1564)	(.1541)
Poor	1.892***	1.922***	1.801***	2.008***	2.020***
	(.1798)	(.1782)	(.1851)	(.1859)	(.1887)
Risk aversion					
Disagree strongly					
Disagree somewhat		-.1720	-.1563	-.1811	-.1773
		(.1334)	(.1334)	(.1296)	(.1296)
Uncertain		-.1853	-.1938	-.1925	-.1883
		(.1606)	(.1591)	(.1545)	(.1538)
Agree somewhat		-.1506	-.1324	-.1556*	-.1343
		(.1401)	(.1408)	(.1334)	(.1301)
Agree strongly		-.7016**	-.6651**	-.5761**	-.5805*
		(.2928)	(.2890)	(.4177)	(.4210)
Cognitive limitations			.3987***	.4606***	.4930***
			(.1350)	(.1255)	(.1273)
Education				.0932***	.0907***
				(.0106)	(.0121)
Poverty Categories					
Poor/negative					
Near poor					-.3191
					(.2755)
Low income					-.0165
					(.1758)
Middle income					.0504
					(.1372)
High income					.0593
					(.1463)

1) Standard error in parentheses

2) Asterisks indicate level of statistical significance: ***p<0.01, **p<0.05, *p<0.

Table 6.5 Log-linear Regression Results of 2010 Data

Log-linear regression with logarithm of total expenditure as the dependent variable

	Reg1	Reg2	Reg3	Reg4	Reg5
Self-rated health					
Excellent					
Very good	.3511**	.3620**	.3668**	.3948***	.4186***
	(.1452)	(.1396)	(.1399)	(.1332)	(.1342)
Good	.6962***	.7210***	.6991***	.8378***	.8660***
	(.1641)	(.1627)	(.1633)	(.1556)	(.1571)
Fair	1.415***	1.455***	1.407***	1.627***	1.680***
	(.1642)	(.1628)	(.1676)	(.1627)	(.1647)
Poor	1.659***	1.728***	1.643***	1.934***	2.005***
	(.2167)	(.2140)	(.2185)	(.2276)	(.2237)
Risk aversion					
Disagree strongly					
Disagree somewhat		-.2296	-.2116	-.2672	-.2577
		(.1729)	(.1724)	(.1722)	(.1719)
uncertain		-.3393**	-.3300*	-.3047**	-.2830*
		(.1633)	(.1622)	(.1534)	(.1547)
Agree somewhat		-.0861	-.0769	-.0962	-.0924
		(.1428)	(.1420)	(.1355)	(.1354)
Agree strongly		-.4226	-.4216	-.2968	-.2438
		(.3648)	(.3648)	(.3407)	(.3507)
Cognitive limitations			.3870***	.5248***	.5294***
			(.1323)	(.1329)	(.1328)
education				.1122***	.1044***
				(.0149)	(.0155)
Poverty Categories					
Poor/negative					
Near poor					-.2272
					(.2932)
Low income					-.0707
					(.2091)
Middle income					.0698
					(.1930)
High income					.2067
					(.1890)

1) Standard error in parentheses

2) Asterisks indicate level of statistical significance: ***p<0.01, **p<0.05, *p<0.1

Table 6.6 Log-linear Regression Results of 2011 Data

Log-linear regression with logarithm total expenditure as the dependent variable

	Reg1	Reg2	Reg3	Reg4	Reg5
Self-rated health					
Excellent					
Very good	.2014	.2060	.2088	.2781	.3058*
	(.1675)	(.1654)	(.1651)	(.1684)	(.1714)
Good	.4938***	.4919***	.4817***	.5937***	.6200***
	(.1479)	(.1481)	(.1454)	(.1447)	(.1514)
Fair	1.160***	1.175***	1.063***	1.239***	1.279***
	(.1626)	(.1573)	(.1600)	(.1597)	(.1655)
Poor	1.995***	2.029***	1.805***	2.046***	2.110***
	(.2005)	(.2001)	(.2008)	(.2022)	(.2158)
Risk aversion					
Disagree strongly					
Disagree somewhat		.0069	.0236	-.0008	-.0016
		(.1694)	(.1701)	(.1716)	(.1709)
uncertain		-.3741*	-.3738*	-.3687*	-.3659*
		(.2128)	(.2103)	(.2046)	(.2042)
Agree somewhat		-.2518*	-.2279	-.2595	-.2624
		(.1611)	(.1607)	(.1590)	(.1599)
Agree strongly		-.5633*	-.5166*	-.4840*	-.4511
		(.2883)	(.2889)	(.2941)	(.2977)
Cognitive limitations			.6813***	.7519***	.7427***
			(.1212)	(.1239)	(.1233)
education				.0878***	.0811***
				(.0146)	(.0145)
Poverty Categories					
Poor/negative					
Near poor					.1600
					(.1771)
Low income					-.0069
					(.1583)
Middle income					.1757
					(.1385)
High income					.2380
					(.1708)

1) Standard error in parentheses

2) Asterisks indicate level of statistical significance: ***p<0.01, **p<0.05, *p<0.1

Table 6.7 Probit Model Results of 2009 Data

Medigap purchase (binary variable) as dependent variable

	Reg1	Reg2	Reg3	Reg4	Reg5
Self-rated health					
Excellent					
Very good	-.3231*	-.3209*	-.3269	-.3105*	-.3129*
	(.1728)	(.1739)	(.1738)	(.1765)	(.1767)
Good	-.2552	-.2627	-.2592	-.1923*	-.1839
	(.1722)	(.1703)	(.1703)	(.1717)	(.1696)
Fair	-.1600	-.1604	-.1262	-.0348	-.0284
	(.1651)	(.1656)	(.1671)	(.1720)	(.1716)
Poor	-.3659	-.3639	-.2747	-.1553	-.1539
	(.2398)	(.2400)	(.2424)	(.2469)	(.2460)
Risk aversion					
Disagree strongly					
Disagree somewhat		.0328	.0213	.0020	-.0001
		(.1474)	(.1473)	(.1452)	(.1466)
uncertain		.1431	.1558	.1485	.1547
		(.1570)	(.1555)	(.1582)	(.1572)
Agree somewhat		.0783	.0720	.0557	.0680
		(.1574)	(.1583)	(.1607)	(.1639)
Agree strongly		-.0309	-.0552	.0557	.0062
		(.2512)	(.2528)	(.2519)	(.2504)
Cognitive limitations			-.3128**	-.2821**	-.2509
			(.1767)	(.1824)	(.1857)
education				.0589***	.0536***
				(.0180)	(.0173)
Poverty Categories					
Poor/negative					
Near poor					-.0763
					(.3051)
Low income					.2191
					(.2325)
Middle income					.2276
					(.2122)
High income					.2560
					(.2183)

1) Standard error in parentheses

2) Asterisks indicate level of statistical significance: ***p<0.01, **p<0.05, *p<0.1

Table 6.8 Probit Model Results of 2010 Data

Medigap purchase (binary variable) as dependent variable

	Reg1	Reg2	Reg3	Reg4	Reg5
Self-rated health					
Excellent					
Very good	-3932**	-.3999**	-.4019**	-.3915**	-.3954**
	(.1868)	(.1849)	(.1845)	(.1853)	(.1874)
Good	-.3484*	-.3483*	-.3334*	-.2906	-.2776
	(.1980)	(.1967)	(.1984)	(.2003)	(.2031)
Fair	-.5484***	-.5589***	-.5169***	-.4521**	-.4177**
	(.1912)	(.1916)	(.1914)	(.1929)	(.1905)
Poor	-.2492	-.2391	-.1637	-.0704	-.0029
	(.2944)	(.2932)	(.2988)	(.3007)	(.2976)
Risk aversion					
Disagree strongly					
Disagree somewhat		-.0725	-.0890	-.1129	-.0756
		(.1619)	(.1616)	(.1628)	(.1642)
uncertain		-.2444	-.2578	-.2608	-.2236
		(.1740)	(.1740)	(.1740)	(.1716)
Agree somewhat		-.1367	-.1504	-.1721	-.1882
		(.2015)	(.2009)	(.1986)	(.2004)
Agree strongly		.0664	.0692	.1157	.2719
		(.3138)	(.3148)	(.3131)	(.3185)
Cognitive limitations			-.3892**	-.3483*	-.3064
			(.2024)	(.2067)	(.2117)
education				.0423***	.0255*
				(.0154)	(.0153)
Poverty Categories					
Poor/negative					
Near poor					.4667
					(.3032)
Low income					.8709***
					(.2388)
Middle income					.7967***
					(.2760)
High income					1.0519***
					(.2621)

1) Standard error in parentheses

2) Asterisks indicate level of statistical significance: ***p<0.01, **p<0.05, *p<0.1

Table 6.9 Probit Model Results of 2011 Data

Medigap purchase (binary variable) as dependent variable

	Reg1	Reg2	Reg3	Reg4	Reg5
Self-rated health					
Excellent					
Very good	.0517	.0916	.0916	.1114	.1536
	(.1918)	(.1790)	(.1784)	(.1748)	(.1725)
Good	-.3192*	-.2777	-.2777	-.2462	-.1871
	(.1866)	(.1727)	(.1726)	(.1695)	(.1704)
Fair	-.6268**	-.6012**	-.6011**	-.5597**	-.5021*
	(.2613)	(.2514)	(.2507)	(.2518)	(.2554)
Poor	-.9042***	-.8743***	-.8742**	-.8104**	-.7405**
	(.3349)	(.3346)	(.3501)	(.3435)	(.3452)
Risk aversion					
Disagree strongly					
Disagree somewhat		.0248	.0248	.0223	.0078
		(.1581)	(.1562)	(.1569)	(.1553)
uncertain		-.1678	-.1678	-.1719	-.1665
		(.2307)	(.2290)	(.2300)	(.2310)
Agree somewhat		.3189	.3189	.3098	.2921
		(.2125)	(.2085)	(.2098)	(.2036)
Agree strongly		.1487	.1487	-.1556	.2125
		(.3042)	(.3036)	(.3056)	(.3152)
Cognitive limitations			-.0003	-.0059	.0101
			(.2249)	(.2266)	(.2283)
education				.0222	.0106
				(.0208)	(.0203)
Poverty Categories					
Poor/negative					
Near poor					.7195***
					(.1993)
Low income					.6854***
					(.2119)
Middle income					.6046**
					(.2401)
High income					.8735***
					(.2285)

1) Standard error in parentheses

2) Asterisks indicate level of statistical significance: ***p<0.01, **p<0.05, *p<0.1

Table 6.10-Table 6.12 show the result from SUR models from 2009 to 2011.

Table 6.10 Seemingly Unrelated Regression of 2009 Data

Dept Var: Logarithm of total health expenditure

	Reg1	Reg2	Reg3	Reg4	Reg5
Self-rated health					
Excellent					
Very Good	.3686***	.3698***	.3750***	.3911***	.3922***
	(.1358)	(1350)	(.1347)	(.1325)	(.1324)
Good	.6576***	.6627***	.6579***	.7799***	.7882***
	(.1330)	(.1322)	(.1314)	(.1305)	(1309)
Fair	1.073***	1.080***	1.024***	1.192***	1.198***
	(.1495)	(.1485)	(.1496)	(.1485)	(.1492)
Poor	1.892***	1.922***	1.801***	2.008***	2.020***
	(.1942)	(.1930)	(.1975)	(.1957)	(.1973)
Risk Aversion					
Disagree Strongly					
Disagree somewhat		-.1720	-.1563	-.1811	-.1773
		(.1228)	(.1227)	(.1207)	(.1207)
Uncertain		-.1853	-.1938	-.1925	-.1883
		(.1316)	(.1314)	(.1292)	(.1291)
Agree somewhat		-.1506	-.1324	-.1556	-.1343
		(.1347)	(.1346)	(.1312)	(.1325)
Agree strongly		-.7016***	-.6651***	-.5761**	-.5805**
		(.2441)	(.2440)	(.2402)	(.2400)
Cognitive limitations			.3987***	.4606***	.4930***
			(.1426)	(.1404)	(.1411)
education				.0932***	.0907***
				(.0110)	(.0115)
Poverty Categories					
Poor/Negative					
Near Pool					-.3191
					(.1958)
Low income					-.0165
					(.1619)
Middle Income					.0504
					(.1498)
High Income					.0593
					(.1527)

1) Standard error in parentheses

2) Asterisks indicate level of statistical significance: ***p<0.01, **p<0.05, *p<0.1

Dept. Var: Medigap purchase (binary variable)

	Reg1	Reg2	Reg3	Reg4	Reg5
Self-rated health					
Excellent					
Very Good	-.3231***	-.3209***	-.3269***	-.3105***	-..3129***
	(.0380)	(.0380)	(.0379)	(.0375)	(.0374)
Good	-.2552***	-.2627***	-.2592***	-.1923***	-.1839***
	(.0372)	(.0372)	(.0371)	(.0370)	(.0370)
Fair	-.1600***	-.1604***	-.1262***	-.0348	-.0284
	(.0418)	(.0418)	(.0421)	(.0421)	(.0422)
Poor	-.3659***	-.3639***	-.2747**	-.1553***	-.1539***
	(.0543)	(.0544)	(.0556)	(.0555)	(.0558)
Risk Aversion					
Disagree Strongly					
Disagree somewhat		-.0328	.0213	.0020	.0001
		(.0346)	(.0345)	(.0342)	(.0341)
Uncertain		.1431***	.1558***	.1485***	.1547***
		(.0371)	(.0370)	(.0366)	(.0365)
Agree somewhat		-.0783**	.0720	.0557	.0680*
		(.0379)	(.0379)	(.0375)	(.0375)
Agree strongly		-.0309	-.0552	.0050	.0062
		(.0688)	(.0678)	(.0681)	(.0679)
Cognitive limitations			-.3128***	-.2821***	-.2509***
			(.0794)	(.0398)	(.0399)
education				.0589***	.0536
				(.0031)	(.0032)
Poverty <u>Categories</u>					
Poor/Negative					
Near Poor					-.0763
					(.0554)
Low income					.2191***
					(.0458)
Middle Income					.2276***
					(.0424)
High Income					.2560***
					(.0432)

1) Standard error in parentheses

2) Asterisks indicate level of statistical significance: ***p<0.01, **p<0.05, *p<0.1

Table 6.11 Seemingly Unrelated Regression of 2010 Data

Dept Var: Logarithm of total health expenditure

	Reg1	Reg2	Reg3	Reg4	Reg5
Self-rated health					
Excellent					
Very Good	.3512**	.3614**	.3664***	.3945***	.4177***
	(.1426)	(.1419)	(.1417)	(.1386)	(.1388)
Good	.6974***	.7217***	.6995***	.8393***	.8667***
	(.1396)	(.1395)	(.1395)	(.1372)	(.1374)
Fair	1.418***	1.455***	1.407***	1.629***	1.681***
	(.1587)	(.1585)	(.1594)	(.1575)	(.1568)
Poor	1.663***	1.730***	1.644***	1.937***	2.007***
	(.2150)	(.2143)	(.2166)	(.2138)	(.2154)
Risk Aversion					
Disagree Strongly					
Disagree somewhat		-.2304*	-.2123*	-.2683**	-.2590**
		(.1287)	(.1287)	(.1177)	(.1259)
Uncertain		-.3576**	-.3484**	-.3240**	-.3027**
		(.1401)	(.1399)	(.1368)	(.1370)
Agree somewhat		-.0878	-.0785	-.0980	-.0943
		(.1435)	(.1433)	(.1401)	(.1400)
Agree strongly		-.4234	-.4223	-.2967	-.2448
		(.2792)	(.2787)	(.1401)	(.2730)
Cognitive limitations			.3907**	.5297***	.5341***
			(.1528)	(.1501)	(.1501)
education				.1130***	.1054***
				(.0116)	(.0120)
Poverty Categories					
Poor/Negative					
Near Poor					-.2227
					(.2163)
Low income					-.0699
					(.1676)
Middle Income					.0696
					(.1555)
High Income					.2017
					(.1574)

1) Standard error in parentheses

2) Asterisks indicate level of statistical significance: ***p<0.01, **p<0.05, *p<0.1

Dept. Var: Medigap purchase (binary variable)

	Reg1	Reg2	Reg3	Reg4	Reg5
Self-rated health					
Excellent					
Very Good	-.3932***	-.3999***	-.4019***	-.3815***	-.3954***
	(.0380)	(.0379)	(.0378)	(.0376)	(.0371)
Good	-.3484***	-.3483***	-.3334***	-.2906***	-.2776***
	(.0372)	(.0373)	(.0372)	(.0372)	(.0367)
Fair	-.5484***	-.5589***	-.5169***	-.4521***	-.4177***
	(.0422)	(.0424)	(.0425)	(.0427)	(.0424)
Poor	-.2492***	-.2391***	-.1637***	-.0704	-.0029
	(.0572)	(.0573)	(.0578)	(.0580)	(.0576)
Risk Aversion					
Disagree Strongly					
Disagree somewhat		-.0725**	-.0890**	-.1129***	-.0756**
		(.0344)	(.0343)	(.0342)	(.0336)
Uncertain		-.2444***	-.2578***	-.2608***	-.2236***
		(.0374)	(.0373)	(.0371)	(.0366)
Agree somewhat		-.1367***	-.1504***	-.1721***	-.1882***
		(.0383)	(.0382)	(.0380)	(.0374)
Agree strongly		.0664	.0692	.1157	.2719***
		(.0746)	(.0744)	(.0740)	(.0730)
Cognitive limitations			-.3892***	-.3483***	-.3064***
			(.0408)	(.0407)	(.0401)
education				.0423***	.0255***
				(.0031)	(.0032)
Poverty Categories					
Poor/Negative					
Near Poor					.4667***
					(.0578)
Low income					.8709***
					(.0448)
Middle Income					.7967***
					(.0416)
High Income					1.0519***
					(.0421)

1) Standard error in parentheses

2) Asterisks indicate level of statistical significance: ***p<0.01, **p<0.05, *p<0.1

Table 6.12 Seemingly Unrelated Regression Results of 2011 Data

Dept Var: Logarithm of total health expenditure

	Reg1	Reg2	Reg3	Reg4	Reg5
Self-rated health					
Excellent					
Very Good	.2014	.2060	.2088	.2781**	.3060**
	(.1389)	(.1383)	(.1376)	(.1361)	(.1367)
Good	.4936***	.4915***	.4814***	.5934***	.6199***
	(.1366)	(.1363)	(.1356)	(.1346)	(.1358)
Fair	1.160***	1.174***	1.0625***	1.283***	1.278***
	(.1574)	(.1569)	(.1580)	(.1576)	(.1591)
Poor	1.994***	2.029***	1.804***	2.045***	2.109***
	(.2066)	(.2064)	(.2110)	(.2107)	(.2128)
Risk Aversion					
Disagree Strongly					
Disagree somewhat		.0067	.0234	-.0011	-.0013
		(.1240)	(.1234)	(.1218)	(.1217)
Uncertain		-.3780***	-.3777***	-.3719***	-.3689***
		(.1348)	(.1341)	(.1323)	(.1323)
Agree somewhat		-.2521*	-.2282*	-.2599*	-.2628*
		(.1432)	(.1425)	(.1407)	(.1407)
Agree strongly		-.5632**	-.5166*	-.4840*	-.4508
		(.2882)	(.2869)	(.2831)	(.2838)
Cognitive limitations			.6814***	.7520***	.7427***
			(.1476)	(.1459)	(.1459)
education				.0878***	.0811***
				(.0117)	(.0122)
Poverty Categories					
Poor/Negative					
Near Poor					.1602
					(.2084)
Low income					-.0081
					(.1680)
Middle Income					.1756
					(.1590)
High Income					.2386
					(.1609)

1) Standard error in parentheses

2) Asterisks indicate level of statistical significance: ***p<0.01, **p<0.05, *p<0.

Dept. Var: Medigap purchase (binary variable)

	Reg1	Reg2	Reg3	Reg4	Reg5
Self-rated health					
Excellent					
Very Good	.0517	.0916**	.0916**	.1114***	.1536**
	(.0378)	(.0378)	(.0378)	(.0749)	(.0376)
Good	-.3192***	-.2777***	-.2777***	-.2462***	-.1871***
	(.0373)	(.0372)	(.0372)	(.0374)	(.0374)
Fair	-.6268***	-.6012***	-.6011***	-.5597***	-.5021***
	(.0430)	(.0429)	(.0434)	(.0438)	(.0438)
Poor	-.9042***	-.8743***	-.8742***	-.8104***	-.7405***
	(.0564)	(.0564)	(.0580)	(.0586)	(.0586)
Risk Aversion					
Disagree Strongly					
Disagree somewhat		.0248	.0248	.0223	.0078
		(.0339)	(.0339)	(.0338)	(.0335)
Uncertain		-.1678***	-.1678***	-.1719***	-.1665***
		(.0368)	(.0368)	(.0368)	(.0364)
Agree somewhat		.3189***	.3189***	.3098***	.2921***
		(.0391)	(.0391)	(.0391)	(.0387)
Agree strongly		.1487*	.1487**	-.1556**	.2125
		(.0788)	(.0788)	(.0787)	(.0781)
Cognitive limitations			-.0003	.0059	.0101
			(.405)	(.0406)	(.0402)
education				.0222***	.0106***
				(.0032)	(.0033)
Poverty Categories					
Poor/Negative					
Near Poor					.7195***
					(.0574)
Low income					.6854***
					(.0463)
Middle Income					.6046***
					(.0438)
High Income					.8736***
					(.0443)

1) Standard error in parentheses

2) Asterisks indicate level of statistical significance: ***p<0.01, **p<0.05, *p<0.1

CHAPTER 7 DISCUSSION AND LIMITATIONS

7.1 Thoughts on "premium variables"

In this study, we control age, gender, and smoking status and assume all of these three variables are used for insurance companies to set their premiums. The issue is that this is not always the case. In some states, the community is used for rating premium, and in other states age rating is used. Community rating premium means that within a certain plan type, all policyholders are charged the same premium regardless of age or health status. For age rating, age is considered and within a plan type policyholders with different ages are charged with different premium. For instance, 80 years old policyholders pay higher premiums than policyholders whose age is 65 even though they are enrolled in same type of plan and live in the same community. Generally speaking, smoking status and gender also have an impact on premium and usually smokers pay higher premium than non-smokers and male pay higher premium than female. However, compared to age, the effect of smoking and gender is less important.

In our paper, we always control age, gender and smoking status (named as "premium variables"). It means that in some extent we overestimate the risk classification assessed by insurance companies and therefore underestimate the selection effect.

7.2 Thoughts on "open enrollment period"

There is a 6 months open enrollment period starting from the first day of the month when the individual turns to age 65. Outside the open enrollment, insurance companies can screen potential beneficiaries based on their pre-existing health condition. Individuals may need to pass the requirements by the insurance companies in order to be able to buy

Medigap insurance. They may charge individuals higher price if they make the purchase during the open enrollment period. Individuals can also be rejected by the insurance companies if they cannot pass the requirements for enrollment. We do not have any variables that can detect whether individuals were enrolled during open enrollment period. There is one possibility that some individuals who would like to enroll in Medigap coverage did not enroll in time during open enrollment period, but submitted their applications outside the period and got rejected by the insurance companies. In other words, cream skimming by the insurance companies cannot be completely ruled out.

7.3 Thoughts on "people age under 65"

Those who are eligible for enrollment in Medicare under age 65 should satisfy at least one of the three following conditions.

- One is that individual receives benefit from Social Security Disabled Insurance or Railroad Retirement Board for at least 24 months and is disabled.
- Second, needing a kidney transplant or continuing dialysis for end stage renal disease (ESRD).
- Third, individual has amyotrophic lateral sclerosis (ALS) and is eligible for Social Security Disability Insurance. Certain plans of Medigap coverage are eligible for these disabled and age under 65 individuals to purchase.

Generally speaking, those who are under age 65 enrolled in Medicare are in worse condition in health compared to others. In our study, we have ruled out this relative special group of individuals.

7.4 Limitations

Medigap market has been standardized in 1992. The basic plan is Plan A and more advanced plans are introduces recently besides Plan A to D, for instance, Plan M and N were effective as new Medigap policies since 2010. Plan M was designed for individuals who would prefer lower monthly premiums for an attractive health insurance plan, but pay higher out of pocket medical expenses in exchange, and Plan N provides co-insurance payments on hospitals costs for Part A, as well as a full year of payments once Medicare benefits have been depleted. Coverage of Medigap varies by different plans. Coverage of advanced plan includes but not limit to what plan covers. Thus, treating Medigap as binary variable do not lose generality but lose part of the detailed information. It will be useful in the analysis if we were able to obtain Medigap types and treat it as a multi-level categorical variable instead.

There are multiple ways to measure realized loss, for instance, total ex post health expenditure, utilization of health care. Using only one-dimensional measurement of total health expenditure could lead to the analysis to some extent vulnerable.

CHAPTER 8 CONCLUSION

In this study, we investigate the selection effect and moral hazard under asymmetric information in the context of Medigap insurance market. We find out the existence of advantageous selection and moral hazard in Medigap insurance market.

We begin by using data of the self-rated health status, the total health expenditure and Medigap purchase conditioned on risk classification used by the insurance companies, along with elaborate controls of more additional variables such as risk aversion, cognitive limitation, years of education, and poverty categories to demonstrate that private information of risk type exists in Medigap insurance market. When we use independent Probit model and log linear regression model, we fail to find positive correlation between private information of risk type and Medigap coverage. When we use SUR model by considering the correlation between the residuals between the total health expenditure and Medigap coverage, we find negative correlation between private information of risk type and Medigap coverage.

We reconcile these findings by presenting evidence of coexistence of moral hazard and advantageous selection in Medigap insurance market.

Our findings also highlight that the frequently examined factor risk aversion does not have consistently effect across years in this study. Individuals with more education are more likely to have slightly higher total health expenditure and slightly more likely to have Medigap coverage. Higher income individuals are more likely to have Medigap coverage, however, not lower in risk type at least in terms of total health expenditure. The only factor, which contributes to advantageous selection, is cognitive limitation.

Therefore, one important direction for subsequent work is to investigate other potential sources that lead to advantageous selection in Medigap insurance market since most of those conventionally focused factors seem to have little effect. Furthermore, it is worth to figure out how advantageous selection would affect efficiency and welfare.

REFERENCES

Abbring, J.H., Chiappori, P.A., Heckman, J.J., and Pinquet, J. (2003). "Moral Hazard and Dynamic Insurance Data." *Journal of the European Economic Association*, Vol. 1, 767-820.

Aperjis, C. and Balestrieri, F. (2011) "Loss Aversion Leading to Advantageous Selection." *HP Laboratories Working Paper*, HPL-2011-209.

Avitabile, C. (2011) "Advantageous Selection in the Private Hospital Insurance Market in Europe: Evidence on the Role of Education and Cognitive Ability." *CSEF Working paper* No. 221.

Bolhaar, J. (2010). "Advantageous Selection in a Dynamic Framework." *Tinbergen Institute Discussion Paper.*

Buchmueller, T.C., Fiebig, D., Jones, G. and Savage, E. (2008). "Advantageous Selection in Private Health Insurance: The Case of Australia." *CHERE Working paper* 2008/2.

Chiappori, P.A., Salanié, B. (2000). "Testing for Asymmetric Information in Insurance Markets." *The Journal of Political Economy*, Vol. 108, No. 1, pp. 56-78

Cohen, A. and Siegelman, P. (2009) "Testing for Adverse Selection in Insurance Markets." *NBER Working Paper* No. 15586.

Cutler, D.M., Finkelstein, A. and McGarry, K. (2008) "Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance." *NBER Working Paper* No. 13746.

Cutler, D.M. and Zeckhauser, R.J. (1998) "Adverse Selection in Health Insurance."

Frontiers in Health Policy Research, Vol. 1, 1-32.

Dardanoni, V. and Donni, P.L. (2012) "Incentive and selection effects of Medigap insurance on inpatient care." *Journal of Health Economics*, 31, 457–470.

de Meza, D. and Webb, D.C. (2001) "Advantageous Selection in Insurance Markets." *The RAND Journal of Economics,* Vol. 32, No. 2, pp. 249-262.

Dempster, A. P., Laird, N. M. and Rubin, D. B (1977) "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society*: *Series B (Methodological),* 39, No 2, 1–22. □

Eichner, M.J. (1998) "The Demand for Medical Care: What People Pay Does Matter." *The American Economic Review: Papers and Proceedings of the Hundred and Tenth Annual Meeting of the American Economic Association*, Vol. 88, No. 2, pp. 117-121.

Einav, L. and Finkelstein, A. (2011) "Selection in Insurance Markets: Theory and Empirics in Pictures." *Journal of Economic Perspectives,* Vol. 25, No. 1, pp. 115-138.

Einav, L., Finkelstein, A., and Ryan, S.P. (2103) "Selection on Moral Hazard in Health Insurance." *American Economic Review,* 2013, 103(1): 178–219.

Fang, H., Keane, M.P., and Silverman D. (2008) "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market." *Journal of Political Economy*, Vol. 116, No. 2, 303-350.

Finkelstein, A., and McGarry, K. (2006). "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market." *American Economic Review*, 96(4): 938-958.

Gao, F., Powers, M.R. and Wang, J. (2009) "Adverse Selection or Advantageous Selection? Risk and Underwriting in China's Health-Insurance Market*." Insurance: Mathematics and Economics*, 44, 505-510.

Hackmann, M.B., Kolstad, J.T. and Kowalski, A.E. (2012) "Health Reform, Health Insurance, and Selection: Estimating Selection into Health Insurance Using the Massachusetts Health Reform." *American Economic Review: Papers & Proceeding*. 102 (3): 498-501.

Handel, B.R. (2013) "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." *American Economic Review*, 103(7): 2643-2682.

Heckman. J. and Singer, B. (1984) "A Method for Minimizing the Impact of Distributional Assumptions in Economics Models for Duration Data." *Econometrica*, Vol. 52, No. 2, pp. 271-320.

Hedengren, D. and Stratmann T. (2012) "Adverse vs. Advantageous Selection in Life Insurance Markets." *SSRN Working Paper*. http://dx.doi.org/10.2139/ssrn.2194376

Hemenway, D. (1990) "Propitious Selection." *The Quarterly Journal of Economics*, Vol. 105, No. 4, pp. 1063-1069.

Huang, R.J., Tzeng, L.Y., and Wang, K.C. (2009). "Empirical Evidence for Advantageous Selection in the Insurance Market*." The Geneva Risk and Insurance Review,* 34, 1-19.

Jeliazkov, I. and Lloro, A. (2011) "Maximum Simulated Likelihood Estimation: Techniques and Applications in Economics*." Computational Optimization, Methods and Algorithms*. Springer, pp 85-100.

Keane, M., and Stavrunova, O. (2011) "Adverse Selection, Moral Hazard and the Demand for Medigap Insurance." *Health, Econometrics and Data Group (HEDG) Working Paper,* 10/14.

Mahdavi, G. (2005) "Advantageous Selection versus Adverse Selection in Life Insurance Market." *Japan Society for the Promotion of Science (JSPS) Working paper.*

Rothschild, M. and Stiglitz, J. (1976) "Equilibrium in Competitive Insurance Markets: an Essay on the Economics of Imperfect Information." *The Quarterly Journal of Economics,* Vol. 90, No. 4, pp. 629-649.

Spenkuch, J.L. (2012) "Moral Hazard and Selection among the Poor: Evidence from a Randomized Experiment." *Journal of Health Economics*, 31, 72–85.

Stanciole, A. (2007) "Health Insurance and Life Style Choices: Identifying the Ex Ante Moral Hazard." *Integrated Research Infrastructure in Social Sciences (IRISS) Working Papers.*

Wolfe, J.R. and Goddeeris, J.H. (1991) "Adverse Selection, Moral Hazard, and Wealth Effects in the Medigap Insurance Market." *Journal of Health Economics*, 10, 433-459.

Yao, Y., Schmit, J.T. and Sydnor, J.R. (2014) "Advantageous or Adverse Selection in Emerging Health Insurance Markets: Evidence from a Micro Health Insurance Program in Pakistan." *Working paper.*

ABSTRACT

SELECTION EFFECT IN MEDIGAP INSURANCE MARKET WITH MULTI-DIMENSIONAL PRIVATE INFORMATION

by

YANG LIU

December 2016

Advisor: Dr. Allen Goodman

Major: Economics

Degree: Doctor of Philosophy

Theoretical models of insurance suggest that when individuals have private information about their risk type alone, insurance coverage will be adversely selected by those riskier individuals due to asymmetric information. In this study we investigate whether individuals have private information of their risk type and riskier individuals are indeed more likely to choose health insurance in the context of Medigap insurance market in the United States where government intervention reinforces information asymmetry. Medigap is supplemental private insurance that optional to those who have Medicare Part A and Part B and it is used to cover some of the cost sharing required by Medicare

We find out that conditional on risk classification assessed by insurance companies, self-rated health status is positively correlated with total health expenditure and negative correlated with Medigap coverage. It demonstrates that people have private information about their risk type; however, those with worse self-rated health status less likely to choose Medigap coverage. This finding provides evidence that on average there is

advantageous selection rather than adverse selection in Medigap insurance market. Meanwhile, we find out coexistence of moral hazard.

The only factor that could be potentially identified as source of advantageous selection in this study is cognitive limitation. Cognitive limitation is positively and significantly correlated with logarithm of total health expenditure in 2009, 2010 and 2011; at the same time, it is negatively and significantly correlated with Medigap purchase regardless of controlling only premium variables or premium variables plus risk aversion, education and poverty categories in Seemingly Unrelated Regression model in 2009 and 2010.

The factors that have been investigated in several previous articles such as education, risk aversion, poverty categories which reflect economic status seem to have little effect if there is any in our study.

AUTOBIOGRAPHICAL STATEMENT

EDUCATION

Ph.D. Economics, Wayne State University, Detroit, MI (expected)12/2016

M.A. Economics, Wayne State University, Detroit, MI 05/2009

B.B.A., China University of Political Science and Law, Beijing, China 07/2005

PROFESSIONAL EXPERIENCE

Data and Policy Analyst, Acumen LLC

PhD Research, Department of Economics, Wayne State University

- Studied the effect of obesity on life expectancy using panel data from nineteen OECD countries spanning over thirteen years using R. Built a statistical regression model on both medical and non-medical factors with fixed effects and random effects. Applied the Hausman test to identify coefficients efficiency.
- Examined selection effect in inpatient care in terms of overnight stay in hospital in Medigap insurance market using HRS (Health Retirement Study) data. Applied Random Forest to impute missing data. Used Logistic regression, Probit model, Bivariate Probit model to test the correlation between Medigap and overnight stay in hospital. STATA and R were used.
- Investigated selection and moral hazard in Medigap insurance market using MEPS (Medical Expenditure Panel Survey) data. SAS and SAS-callable SUDAAN are used in order to deal with complex survey design with weight, stratification and clustering.
-

Instructor, Department of Economics, Wayne State University

- Ten semesters independent instructor for courses including Macroeconomics, Microeconomics and Principles of Economics starting from 2009.

SKILLS

Basic statistical analysis (GLM including Linear Model, Logistical Model, Mixed effect Model etc. Ridge regression, LASSO, Nonparametric Smoothing estimation). Machine learning skills including supervised and unsupervised methods (SVM, Random Forest, K-means, Hierarchical clustering, K-NN etc.)

Computer Languages: STATA, R, SAS with base and advanced certificates, SUDAAN, SQL.